

U.S. Department of Energy

Office of Science; Office of Biological and Environmental Research; Recommendations for Sequencing Targets in Support of the Science Missions of the Office of Biological and Environmental Research (BER)

AGENCY: U.S. Department of Energy (DOE).

ACTION: Notice of Recommendations for Sequencing Targets.

SUMMARY: This Federal Register Notice seeks the input and nominations of interested parties for candidate microbes, microbial consortia, and 100Mb-or-less-sized organisms for draft genomic sequencing in support of Office of Biological and Environmental Research (BER) programs, among them, the Climate Change Research Program, the Natural and Accelerated Bioremediation Research (NABIR) Program, the Environmental Management Science Program (EMSP), the Microbial Genome Program (MGP), and the GTL Program. Nominated candidates should be relevant to DOE mission needs, e.g., organisms involved in environmental processes, including waste remediation, carbon management, and energy production. This announcement is not an offer of direct financial support for research on these organisms. Those nominations selected will result in the DNA sequence of selected organisms being determined at a draft level (6-8 x coverage) at the DOE Production Genomics Facility (PGF) at the Joint Genome Institute (JGI), (<http://www.jgi.doe.gov>). This announcement is designed to assist DOE in determining and prioritizing a list of microbes, microbial consortia, or modest-genome sized (not more than 100Mb) organisms (including eukaryotes) that address DOE mission needs. Following merit review, and subject to the availability of funding and programmatic relevance, draft sequencing will be carried out at the PGF.

DATES: To assure consideration, nominations in response to this notice should be received by 4:30 p.m. (EST), June 26, 2003, to be accepted for merit review. It is anticipated that review will be completed early in the fall of 2003 with draft sequencing at the DOE PGF to commence towards the end of 2003 or early in 2004, conditional upon the provision of high quality DNA.

ADDRESSES: Nominations responding to this notice should be sent to Dr. Daniel W. Drell, Office of Biological and Environmental Research, SC-72, Office of Science, U.S. Department of Energy, 1000 Independence Ave., SW, Washington, DC 20585-1290; email is acceptable and encouraged for submitting nominations using the following addresses: joanne.corcoran@science.doe.gov and daniel.drell@science.doe.gov.

FOR FURTHER INFORMATION CONTACT: Dr. Daniel W. Drell, SC-72, Office of Biological and Environmental Research, Office of Science, U.S. Department of Energy, 1000 Independence Ave. SW, Washington, DC 20585-1290, phone: (301) 903-4742, email: daniel.drell@science.doe.gov. The full text of this notice is available via the Internet using the following Web site address:
<http://www.sc.doe.gov/ober/microbial.html>.

SUPPLEMENTARY INFORMATION: The DOE Office of Biological and Environmental Research supports fundamental research in a variety of missions

(http://www.sc.doe.gov/ober/ober_top.html). Relevant BER programs may include the Climate Change Research Program, the Natural and Accelerated Bioremediation Research (NABIR) Program, the Environmental Management Science Program (EMSP), the Microbial Genome Program (MGP) and the GTL program. The Climate Change Research Program supports investigations of microbially-mediated fixation of atmospheric CO₂. The NABIR program supports research on microbial biotransformations and/or immobilization of metal and radionuclide wastes. The EMSP supports research into microbially-mediated biotransformations of DOE-relevant organic wastes such as chlorinated solvents. The MGP supports key DOE missions by providing and analyzing microbial DNA sequence information to further the understanding and application of microbiology relating to energy production, chemical and materials production, environmental carbon management, and environmental cleanup. The GTL Program builds on the successes of the DOE Human Genome Program (HGP) by seeking to understand biological function in DOE mission relevant microbes with emphases on identifying the multi-component protein complexes in cells, characterizing gene regulatory networks, probing the functional capabilities of the environmental microbial repertoire of genes, and beginning to model these processes computationally.

Over the last nine years, sequencing of a range of microorganisms that live in a wide diversity of environments has provided a considerable information base for scientific research related not only to DOE missions, but also to other federal agency missions and U.S. industry. (<http://www.tigr.org/tdb/mdb/mdbcomplete.html>, <http://www.ornl.gov/microbialgenomes/organisms.html> and http://www.jgi.doe.gov/JGI_microbial/html/). Nonetheless, most of our current knowledge of microbiology still is derived from individual species that either cause disease or grow easily and readily as monocultures under laboratory conditions and are thus easy to study. The preponderance of species in the environment remains largely unknown to science. Many are thought to grow as part of interdependent consortia in which one species supplies a nutrient necessary for the growth of another. Virtually nothing is known of the organization, membership, or functioning of these consortia, especially those involved in environmental processes of DOE interest. Fungi and small multicellular eukaryotes play important roles in the environment as well.

Genomic analyses of sequenced microbes have suggested that processes such as lateral gene transfers at various times in the evolutionary history of some microbial lineages may have blurred the understanding of their phylogenetic relationships. For this notice, groups of microbes that may have exchanged (or may be exchanging) genetic information via lateral gene exchange or plasmid mediated exchanges can be proposed if the processes of genetic exchange result in functions relevant to DOE missions noted above.

Genomic analyses are badly needed of microbial consortia and species refractory to laboratory culture that play important roles in environments challenged with metals, radionuclides, chlorinated solvents, or are involved in carbon sequestration. The candidate(s) must mediate or catalyze metabolic events of energy or environmental importance. Priority will be given to studies on those microbes or microbial consortia that can bioremediate metals and radionuclides, degrade significant biopolymers such as celluloses and lignins, produce potentially useful energy-related materials (H₂, CH₄, ethanol, etc.), or fix or sequester CO₂.

For this notice, candidate organisms (either individual organisms, consortia of organisms, or eukaryotes with small genomes) comprised of archaea, bacteria, fungi, algae, and other eukaryotes with genome sizes not greater than 100Mbp can be proposed for draft sequencing. For a current list of microbes that have been and are being sequenced see <http://www.ornl.gov/microbialgenomes/organisms.html> and <http://www.ornl.gov/microbialgenomes/seq2003.html>.

Aims: This request for nominations of candidate sequencing targets has two broad foci:

(1) Single culturable organisms. These may be bacteria, archaea, fungi, microalgae or multicellular organisms with genomes not larger than 100Mb. The criteria that will be used to evaluate proposed candidates for draft sequencing will include:

- (a) The candidate has significant relevance to the DOE missions noted above;
- (b) The genome size and structure are known;
- (c) The source of genomic DNA (i.e., strain or isolate, and researcher) is identified, and a clonal sample (or samples with low and characterized polymorphism) are available;
- (d) A brief description of post sequencing follow-up work (e.g., a data use plan and how will data be annotated to enable rapid and open use) is included;
- (e) The availability of a DNA/gene transfer system supporting genetic analyses is known;
- (f) Biological novelty or uniqueness (i.e., unusual genetically determined characteristics pertinent to DOE missions) is described;
- (g) Place in the currently understood, 16s RNA based, “Tree of Life” is identified, e.g., is the proposed candidate in a sparsely populated or more heavily populated section of the tree?
- (h) A brief description of the user community is given;
- (i) The potential impact on the scientific community is large;
- (j) Explicit commitment to a data-release schedule, consistent with the guidelines given below is provided.

(2) Currently unculturable or hard-to-culture organisms and environmental consortia. The review criteria that will be used to evaluate proposed candidates for draft sequencing will include most of the criteria listed above for single culturable organisms (with less emphasis on genome size/structure, presence/absence of a genetic system, or position in the “Tree of Life” since it is recognized that few data on these attributes will be available), but in addition, the following considerations will be included:

- (a) Some measure of the “complexity” of the target consortium being proposed, e.g., approximate number of species, size(s) of genomes, and proportions of different members (it is understood that in most cases, only estimates of these parameters may be available) is discussed. When the environmental consortia are complex, approaches should be described to normalize the DNA libraries in order to reduce the amount of sequencing required and assure adequate sampling of the complexity of the consortia. Additionally, the proposer(s) should be prepared to work together with JGI scientists to optimize the yield from the sequencing effort required;
- (b) Past attempts to cultivate consortium members are described, e.g., have any members of this consortium been successfully cultured;

(c) Some spatial/temporal/hydrochemical/geochemical or other characterization of the environment is given, e.g., the physicochemical parameters of the site from which the selected community is derived; a description of the site contaminants; the accessibility of the site for future sampling; the adequacy of site documentation;

(d) If proposed, technical approaches and technology development specific for defining and isolating members of a given consortium are described;

(e) Some indication of the biological function of the relationships, within consortium members, where available along with a discussion of the scientific and programmatic importance of understanding these relationships is given;

(f) Information where available is given about the phylogenetic position(s) of the members of the consortium and what is known about closely related organisms.

(g) Available informatics tools and annotation plan (e.g., for annotating genes from a consortium analysis or grouping identified genes into a putative “consortium phenotype” within the chosen environment) are described;

(h) Explicit commitment to a data-release schedule, consistent with the guidelines given below is provided.

Scientific community standards regarding access to sequencing data are evolving. BER conforms to the general guidance contained within the Draft Rapid Data Release Policy (<http://www.genome.gov/page.cfm?pageID=10506537>) for “community resource projects.” The usual and customary practice for the JGI is to put all sequencing data up on its web site (<http://www.jgi.doe.gov/>) at frequent and periodic intervals. However, for the purposes of this notice, BER does not regard individual genome sequencing efforts involving less than 50Mb, or microbial community sequencing efforts, as requested herein, as “community resource projects” within the definition of the Draft Rapid Data Release policy. BER’s position, which is provisional and subject to evolution, is that no more than 6 months from the completion of 6 x coverage of the sequence for a single-genome project, the data will be released on the JGI web site or to a publicly accessible database with no use restrictions. For microbial community projects, the JGI will conduct normal QA/QC assessments on the sequence output (at approximately 2 x coverage), then discuss with the proposer(s) and with BER staff the extent to which sequencing will be continued to achieve a satisfactory genomic “view” of the selected microbial community. From the time of initiation of this discussion, not more than 6 months will be permitted to elapse before unconditional release of these data. BER is fully aware that some ambiguity remains in the precise initiation of this “embargo” period but stresses its intention and commitment to the rapid release, without any use restrictions, of this data into publicly accessible databases.

Upon selection of a nominated microbial sequencing target, BER expects that Principal Investigators will collaborate with the JGI by providing high quality, high MW genomic DNA for library construction as well as assisting in annotating the draft sequence data until a sufficiently complete annotation is achieved (understanding that this will be sensitive to hypothetical gene predictions and the usual uncertainties of annotation). Following data acquisition and annotation, DOE expects that those whose nominations have been sequenced will make good faith efforts to publish in the open scientific literature the results of their subsequent work, including both the genome sequences of

the organisms sequenced under this notice as well as the annotation. (BER also expects the Principal Investigator of a selected effort to either deposit a culture of the microbe or consortium into a publicly accessible collection or repository, or make it available directly so others can have access). These parties are encouraged to create process- and cost-effective partnerships that will maximize data production and analysis, data dissemination, and progress towards understanding basic biological mechanisms that can further the aims of this effort. Additionally, it must be explicitly understood that DOE will provide an assembled and computationally annotated draft (roughly 6 x; carried out in a paired-end sequencing approach) sequence of the microbe(s) selected, but that research using that sequence data should be funded from separate sources and/or separate solicitations. Finally, there is no commitment to finish a given drafted sequence, although this option may be considered at a later time.

Submission Information: Interested parties should submit a brief white paper to either of the foci given above, consisting of not more than 5 pages of narrative exclusive of attachments (which are discouraged) responding to each of the specific criteria set forth. Electronic submission (to joanne.corcoran@science.doe.gov and daniel.drell@science.doe.gov) is strongly encouraged. It is expected that the Principal Investigator will serve as the main point of contact for additional information on the nominated microbe. Nominations must contain a very short abstract or project summary and a cover page with the name of the applicant, mailing address, phone, fax, and email. The nomination should include 2-page curriculum vitae of the key investigators; letters of intent from collaborators (suggesting the size of the interested community) are permitted.

Nominations will be reviewed relative to the scope and research needs of the BER relevant programs cited above. A brief response to each nomination will be provided electronically following merit and programmatic reviews.

Other useful web sites include:

DOE JGI Microbial Sequencing Priorities for FY2003:

<http://www.ornl.gov/microbialgenomes/seq2003.html>

Microbial Genome Program Home Page--<http://www.sc.doe.gov/ober/microbial.html>

DOE Joint Genome Institute Microbial Web Page--

http://www.jgi.doe.gov/JGI_microbial/html/

GenBank Home Page--<http://www.ncbi.nlm.nih.gov/>

Human Genome Home Page--<http://www.ornl.gov/hgmis>

DOE Genomes to Life--<http://DOEGenomestoLife.org>

DOE Natural and Accelerated Bioremediation Research (NABIR)

Program--<http://www.lbl.gov/nabir>

Biotechnology Investigations--Ocean Margins Program--

<http://www.sc.doe.gov/ober/GC/omp.html>